Distributed Learning for Localization in Wireless Sensor Networks

MinglinYao

Department of Information Engineering, Tangshan, Hebei, 063000, China

yaominglin73@126.com

Abstract

The problem of distributed or decentralized detection and estimation in application such as wireless sensor networks has often been considered in the framework of parametric models, in which strong assumptions are made aout a statistical description of nature. So the distributed learning method is borrowed to solve the localization problem. It assumes that a network with a number of beacon nodes that have perfect knowledge of its own coordinates and utilizes their knowledge as training data to perform the above classification. In this thesis, three approaches for distributed learning based on the different features that is used to determine the class of each node have been proposed, namely, the hop-count (HC) method, the density-aware hop-count length (DHL) method, and the distance vector (DV) method. These methods are compared under different system parameters and also compared with the triangulation method that is often employed in the literature. The simulation results show that the localization methods based on the distributed learning is more accurate and effectual.

Keywords

Distributed Learning, Wireless Sensor Networks; Hop-distance Estimation, Hop-count (HC) Method; Density-aware Hop-count Length (DHL) Method; Distance Vector (DV) Method

Introduction

Wireless sensor networks (WSNs) typically consist of a large number of sensors deployed over a wide area to retrieve information from the environment. Due to the large scale deployment and the limitations in size and cost, sensors' communications are often strictly constrained in terms of energy and bandwidth resources. It is important to utilize the resources efficiently in terms of transmitting only the data that is relevant in achieving the system task, instead of transmitting raw observations.

Many sensor network applications are based on the fundamental task of decentralized detection or estimation where optimal local quantization or decision schemes are derived under constraints on the communication channel. However, most of these works assume parametric models that rely on perfect knowledge of the statistical observation models. these requirements are difficult to obtain in practice, especially as the range of applications or as the target environment widely varies. In localization applications, the environment may vastly differ from one application to another and the prior knowledge or accurate modelling of the environment may not be available and the distributed inference techniques derived under specific parametric models may not be robust or may even be inapplicable.

Due to these reasons, several recent works have taken the nonparametric approach in studying the decentralized detection and estimation problems in sensor networks. In particular, the series of work taken on the learning-theoretic approach have received considerably attention and will be the focus of this thesis. In the literature on statistical learning, a set of training data is assumed available at the decision center, where computation is made by comparing new observations with that set of data. But, in WSN applications, training data may not be available at a single fusion center but instead distributed among sensors. The task in distributed learning is then to communicate useful information regarding the observations as well as the training data to perform the distributed inference task under communication constrains.

In the distributed learning scheme proposed in, training data is distributed among different locations and, if each time a user, or fusion center, desires an estimation or decision, it will first transmit its local observation to sensors that contain the training data. Then theses sensors will compute a binary response based on the received observation and their local training data, and pass it back to the user.

In this design, a WSN is considered that consists of a set of reference nodes that has knowledge of its own location as well as man oblivious sensors that can only estimate its location based on locally obtained

information. More specially, the sensor field is divided into multiple partitions and the distributed learning algorithm is used to determine which partition the target sensor falls into. The location of the reference nodes serve as training data is used to compute this information.

Localization in WSNs have been studied extensively in the past, and most work in the literature is focused on range-based schemes that require a form of estimate on the distance between two or more sensors. Methods to obtain the distance estimate are such as RSSI, TOA, TDOA and AOA, in these methods, every sensor must be equipped with associated ranging hardware which may not be achievable in practice. Hence in this work, we consider the range-free alternative where distributed learning for localization is performed with only local hop-count information between the target sensor and the reference nodes.

Two approaches have been taken into account: the approach where hop-count information is used directly as input to the learning algorithm and the approach where hop-count information is first converted to distance estimates before being used.

Distributed Learing in Wireless Sensor Networks

Consider a network of n sensors denoted by S1, S2, ...,Sn, with direct transmission links to the fusin center in Fig. 1. Suppose that each sensor, say Si, contains a training data (xi, yi), where $y_i \in R$ is the true value of the event and $x_i \in R^d$ is the observation corresponding to the event. In the literature on statistical learning, xi's are known as features, observations, or inputs, and yi's are known as labels or outputs. Here the fusion center represents the user that is to perform the detection or estimation task. Specifically, the fusion center first makes an observation X and broadcasts it to all sensors. Upon reception of the observation from the fusion center, each sensor will choose whether or not to respond to the fusion center's request for information. Assume that each sensor may transmit at most 1 bit to the fusion center. If it choose to respond, a local binary decision, $\delta_{ni} \in \{-1,1\}$, based on the available information, (X, xi, yi), will be computed and sent to the fusion center; otherwise, it will abstain from transmitting. At the fusion center, the local decisions are combined to compute the final decision $g_n\left(X, \{\delta_{ni}\}_{i=1}^n\right)$.

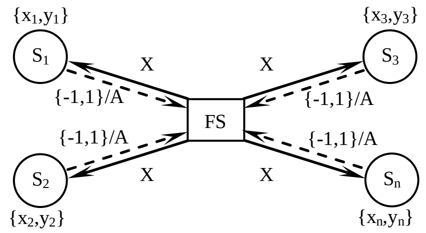


FIG.1 SYSTEM MODEL

To be consistent with the remainder of this thesis, let us consider a binary classification problem where the label is binary, $Y \in \{-1,1\}$. The local decision at sensor Si is given by

$$\delta_{ni} = \begin{cases} y_i & if\ x_i \in B_{RN}(X) \\ abstain & otherwise \end{cases} \tag{1}$$

where $B_{rn}(x) = \{x' \in R^d : \|x - x'\| \le r_n\}$. The final decision at the fusion center is based on the majority vote rule:

$$g_n(X) = \begin{cases} 1 & if\ \dfrac{\sum_{i \in I_V} \delta_{ni}}{|I_V|} \ge 0 \\ -1 & otherwise \end{cases} \tag{2}$$

where Iv is the set of sensors that do not abstain from transmission, $I_V = \{i \in \{1,..., n\} : \delta_{ni} \ne abstain\}$.

Suppose that Y is the true event or label corresponding to the observation X and let $\ell\left(g_n(X), Y; \{x_i, y_i\}_{i=1}^n\right)$ be the loss of the decision rule given the set of training data $\{x_i, y_i\}_{i=1}^n$. For the binary classification problem considered above, the loss is defined as $\ell\left(g_n(X), Y\right) = 1_{\{g_n(X) \ne Y\}}$. Furthermore, let us define L* as the error probability obtained with the optimal Bayesian detection where the joint statistics of (X,Y) are assumed to be perfectly known.

Definition 1: Let $L_N = E\left[\ell(g_n(X), Y)\right]$. Then, $\{g_n\}_{n=1}^\infty$ is said to be universally consistent if $L_n \to L^*$ for all joint distributions of (X,Y).

It has been shown in that the above decision rules are sufficient to achieve universal consistency under certain conditions provided *b* the Stone's theorem, which is stated below.

Theorem 1. If $r_n \to 0$ and $r_n^d \to \infty$ as $n \to \infty$, then

$E[L_N] \to L^*$ for all joint distributions of (X,Y).

The results on universal consistency provided in [8] are proven for the case without errors on the channels. However, in practical systems, noise always exists in the transmission channels, so, we are going to consider this problem.

Assume that there is error existing on the transmission channel from the sensor *i* to the fusion center. In particular, suppose that, the fusion center will receive the true local decision of sensor Si if $\varepsilon_i = 0$ and the wrong decision if $\varepsilon_i = 1$. Thus the new received decision from sensor Si is $\delta_{ni}^{'} = (-1)^{\varepsilon_i}\delta_{ni}$. The decision made at the fusion center then becomes

$$g_n^{'}(X) = \begin{cases} 1 & if\ \frac{\sum_{i\in I_V}\delta_{ni}^{'}}{|I_V|} \geq 0 \\ -1 & otherwise \end{cases} \tag{3}$$

Therefore, it has

$$\frac{\sum_{i\in I_V}\delta_{ni}^{'}}{|I_V|} = \frac{\sum_{i\in I_V \cap correct}\delta_{ni} - \sum_{i\in I_V \cap error}\delta_{ni}}{|I_V|} = \frac{\sum_{i\in I_V}\delta_{ni} - 2\sum_{i\in I_V \cap error}\delta_{ni}}{|I_V|} \tag{4}$$

and also, as N goes to infinity, it has

$$\frac{\sum_{i\in I_V}\delta_{ni}}{|I_V|} + 2P_e > \frac{\sum_{i\in I_V}\delta_{ni}^{'}}{|I_V|} > \frac{\sum_{i\in I_V}\delta_{ni}}{|I_V|} - 2P_e \tag{5}$$

where P_e is the error probability over the network. From the above, we guaranteed that the system with transmission errors yields the same decisions as the errorless case if $\frac{\sum_{i\in I_V}\delta_{ni}}{|I_V|} > 2P_e$ or if $\frac{\sum_{i\in I_V}\delta_{ni}}{|I_V|} < -2P_e$.

Distributed Localization Methords

Distributed Localiz Ation Using Hop-Counts (HC)

Let's consider a network of n nodes S1,…,Sn deployed over a square area [0,D]×[0,D](D>0). Among the n sensors, k of them serve as beacon nodes which have knowledge of their exact geographical coordinates.

Let (x(Si),y(Si)) be the true coordinates of node Si. Based on the sensors location in the x-dimension, and the sensors are divided into M-1 disjoint classes Cx,1,…,Cx,M-1 where $S_i \in C_{x,t}$ if $x(S_i) \geq tD/M$ the position of sensor Si is on the right side of the line x=tD/M, for t=1,2,…,M-1. So, it defines:

$$\alpha_{it}^{x} = \begin{cases} 1 & if\ S_i \in C_{x,t} \\ -1 & if\ S_i \in C_{x,t}^{c} \end{cases} = sign(x(S_i) - tD/M) \tag{6}$$

where $C_{x,t}^{c}$ is the complement set of $C_{x,t}$. Similarly, we can also partition the sensors into M-1 disjoint classes Cy,1,…,Cy,M-1 based on their location in the y-dimension, where $S_i \in C_{y,t}$ if $y(S_i) \geq tD/M$. So,

$$\alpha_{it}^{y} = \begin{cases} 1 & if\ S_i \in C_{y,t} \\ -1 & if\ S_i \in C_{y,t}^{c} \end{cases} = sign(y(S_i) - tD/M) \tag{7}$$

where $C_{y,t}^{c}$ is complement set of $C_{y,t}$. The class $C_{y,t}$ contains nodes that lie above the horizontal line y=tD/M.

By means of the techniques derived from the distributed learning literature, the goal is to determine which classes the target sensor belongs to. This effectively localizes the sensor within D/M precision on each dimension. The localization algorithm with distributed learning algorithm is described below:

1) Initialization Phase: Upon deployment of sensors, each beacon node will first broadcast a HELLO message to all sensors in the network. The HELLO message contains a "hop-count" entry that will be increased by 1 and each time an additional hop is taken. If a redundant HELLO message is received at a sensor, the one with the larger hop-count is discarded. With this initialization, each node, or node Si will be able to obtain its hop-count vector hi. The beacon nodes will send this HELLO message periodically to be adaptive to dynamically varying environments.

2) Local Localization Phase: When a non-beacon node desires an estimate of its location, it will first broadcast an INFO message containing both its hop-count vector and its ID to all beacon nodes. Based on the hop-count vector, each beacon will then compute a local binary classification on each of the 2M-2 classes, namely, Cx,1,…,Cx,M-1 and Cy,1,…,Cy,M-1. Specially, corresponding to the INFO message sent by sensor Si, beacon node Sj will generate a response for class Cx,t as:

$$\delta_{ijt}^{x} = \begin{cases} \alpha_{jt}^{x} & if\ h_{i,j} \le \left\lceil \dfrac{\left|x(S_j) - tD/M\right|}{r} \right\rceil \\ abstain & otherwise \end{cases} \tag{8}$$

where $\lceil a \rceil$ is the smallest integer such that $\lceil a \rceil \ge a$, hi,j is the hop-count from sensor Si to beacon Sj, and similarly for the classes on the y-dimension.

3) Final Decision Phase: Upon reception of the local decision vectors, sensor Si can then use the majority vote fusion rule, similar to that in equation (2), to perform the final decision on each class. Specifically, the final decision on class Cx,t is made by taking

$$g_{i,t}^{x} = \begin{cases} 1 & if\ \sum_{j \in I_V} \delta_{ift}^{x} \ge 0 \\ -1 & otherwise \end{cases} \tag{9}$$

Where Iv is the set of beacon nodes that do not abstain from transmission. Here, $g_{i,t}^{x} = 1$ indicates that Si belongs to the class Cx,t and $g_{i,t}^{x} = -1$ indicates that it belongs to the compliment class $C_{x,t}^{c}$. Similarly, the decision on the class Cy,t is

$$g_{i,t}^{y} = \begin{cases} 1 & if\ \sum_{j \in I_V} \delta_{ift}^{y} \ge 0 \\ -1 & otherwise \end{cases} \tag{10}$$

Given the decision vector $g_i^x = \lfloor g_{i,1}^x, \ldots, g_{i,M-1}^x \rfloor$, the localization on the x-dimension can be performed by traversing the binary decision tree, as shown in Fig.2.

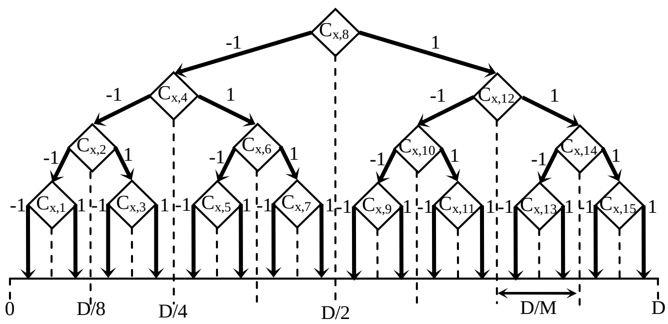


FIG.2 DECISION TREE

Where M=16. Specifically, starting from the root class Cx,t, where t=M/2, sensor Si will go on to decide on class Cx,tl (where tl denotes the left child of the current class in the decision tree) if the decision gi,t=-1. Otherwise, it will move on to the class Cx,tr (where tr denotes the right child of the current class in the decision tree) if the decision gi,t=1. The pseudo-code is given as follows (the x-dimension localization):

① t=M/2 (start at root of the tree Cx,M/2);

② IF (this sensor predicts S not in class Cx,t)

a) IF (Cx,t is a leaf node) THEN RETURN $\hat{x}(S) = (t - 1/2)D/M$

b) ELSE Move to leftchild Cx,tl and set t=tl

③ ELSE

a) IF (Cx,t is a leaf node) THEN RETURN $\hat{x}(S) = (t + 1/2)D/M$

b) ELSE Move to rightchild Cx,tr and set t=tr

④ GOTO Step 2

A similar procedure is also performed to determine the location of each sensor on the y-dimension. The estimated location coordinates for the non-beacon node Si is thus $(\hat{x}(S_i), \hat{y}(S_i))$.

Distributed Localization Using Estimated Hop-Distances

The localization which performed by utilizing the hop-count information is simple and practical, however, the disadvantage is that, as the transmission radius increases or as the density of the network increases, the hop-count cannot accurately reflect the actual distance between nodes and, thus, affecting the accuracy of the location estimate. Although many ranging techniques such as RSSI, TOA, TDOA and AOA can provide explicit estimates of inter-node distances which may allow us to overcome the disadvantages of using only hop-count information, the additional hardware required may not be available, especially for heterogeneous or generic sensor networks.

Recently, several authors proposed methods to estimate hop-distances based only on hop-count parameters, such as the density-aware hop-count localization (DHL) method and distance-vector (DV) method. Usually, both of DHL method and DV method use triangulation to estimate the distance between two nodes in the localization process. Here, instead, the distance estimates with obtained DHL and DV as inputs to the distributed learning algorithm will be utilized.

Suppose that d(i, j) is the estimated distance from the non-beacon node Si to beacon node Sj obtained by either the DHL or the DV method. The response at beacon node Sj is as follows as

$$\delta_{ijt}^{x}=\begin{cases}\alpha_{jt}^{x} & \text{if } d(i,j)\le\left|x(S_j)-tD/M\right|\\ \text{abstain} & \text{otherwise}\end{cases} \tag{11}$$

on the x-dimension, and similarly as

$$\delta_{ijt}^{y}=\begin{cases}\alpha_{jt}^{y} & \text{if } d(i,j)\le\left|y(S_j)-tD/M\right|\\ \text{abstain} & \text{otherwise}\end{cases} \tag{12}$$

on the y-dimension. The final decision is again given by equation (9) and (10).

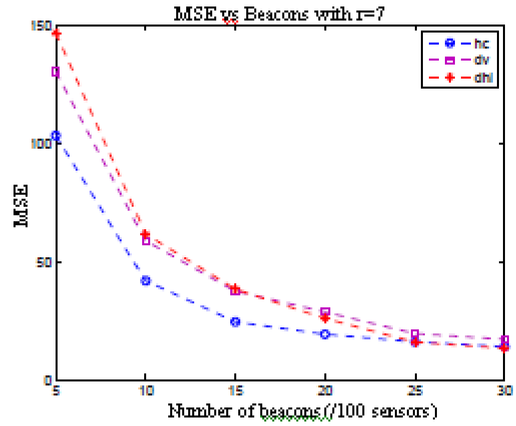
Simulation Results

In the experiment, assume that n=100 sensors are randomly deployed in a square region of [0,D]×[0,D], where D is set to be 30 m. And the transmission range is set to be r=7,10 m. In Fig.3, it compares the performance of the HC method, DV method, and DHL method,, it shows that, for small r (r=7), the HC method more accurately reflects the distance to the destination. Also, the increase in hop-counts provides a more unique signature for nodes that are at different locations. When the number of beacons is small, the DV and DHL methods cannot accurately estimated the hop-count length and, thus the HC method performs better as well. However, as the number of beacon nodes increases for r=10, the DV and DHL methods outperform the HC method since the hop-count no longer accurately reflects the traversed distance and also the decrease in hop-count number makes the hop-count vector less identifiable for sensors at different locations as well.

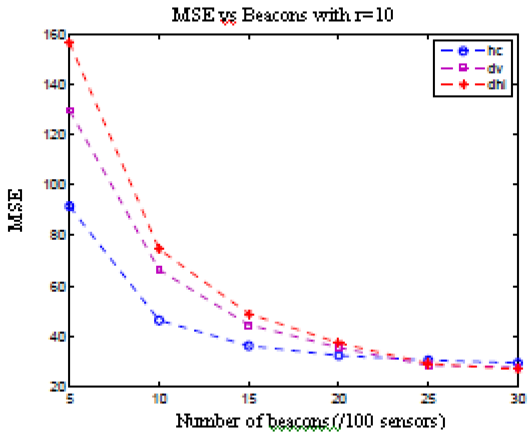
The advantage of DV and DHL is even more clearer as the node density increase as shown in Fig.4. Here, the methods for r=8 are compared in a fixed area with increasing node density. In particular, the number of sensors is increased from 100 to 300 while the number of beacon nodes is fixed at 30; It shows that even though the mean square error (MSE) of all schemes increases with the node density, the loss with hop-distance estimation is smaller and thus it eventually outperforms the hop-count method. This is because, for networks with high node density, both DHL and DV methods will more accurately reflect the actual distance between nodes than the hop-count method.

DHL and DV Methods Using Triangulation

The DHL and DV localization methods utilize the distance estimates obtained from the above methods to perform triangulation. Assume that there are k beacon nodes in the network, denoted by S1, S2, …, SK, at locations (x1, y1), (x2, y2),…, (xk, yk). Let (x0,y0) be the true location of the non-beacon node. Suppose that the non-beacon node obtains the estimated hop-distance to



(a) r=7m



(b) r=10m

FIG.3 COMPARISION ON THE HC, DV, AND DHL METHOD

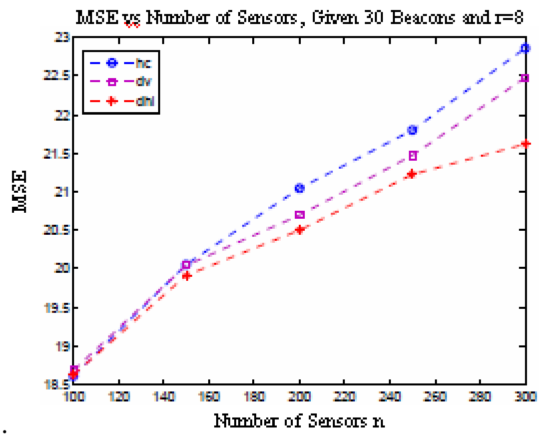


FIG.4 NODE DENSITY INCREASING COMPARISON OF HC, DVAND DHL WITH R=8

each beacon node as $\hat{d}_{0i}$, for i=1,…, k, using either the DHL or the DV method. The error of the hop-distance

estimate $\hat{d}_{0i}$ is given by

$$\varepsilon_{0i} = \hat{d}_{0i} - \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} \tag{13}$$

If we set the error to zero and rearranged the terms, it will has

$$(x_0^2 + y_0^2) - 2x_0 x_i - 2y_0 y_i = \hat{d}_{0i}^2 - x_i^2 - y_i^2 \tag{14}$$

for i=1,..., k. By subtracting each equation with the equation corresponding to i=k, we get rid of the quadratic terms and yield the following equations:

$$\begin{aligned} 2x_0(x_k - x_1) + 2y_0(y_k - y_1) &= \hat{d}_{01}^2 - \hat{d}_{0k}^2 \\ &\quad - x_1^2 - y_1^2 + x_k^2 + y_k^2 \\ 2x_0(x_k - x_{k-1}) + 2y_0(y_k - y_{k-1}) &= \hat{d}_{0,k-1}^2 - \hat{d}_{0k}^2 \\ &\quad - x_{k-1}^2 - y_{k-1}^2 + x_k^2 + y_k^2 \end{aligned} \tag{15}$$

We can rewrite the equations in the form Ax=b where

$$A = \begin{bmatrix} 2(x_k - x_1) & 2(y_k - y_1) \\ 2(x_k - x_2) & 2(y_k - y_2) \\ \cdots & \cdots \\ 2(x_k - x_{k-1}) & 2(y_k - y_{k-1}) \end{bmatrix}$$

$$b = \begin{bmatrix} \hat{d}_{01}^2 - \hat{d}_{0k}^2 - x_1^2 - y_1^2 + x_k^2 + y_k^2 \\ \hat{d}_{02}^2 - \hat{d}_{0k}^2 - x_2^2 - y_2^2 + x_k^2 + y_k^2 \\ \hat{d}_{0,k-1}^2 - \hat{d}_{0k}^2 - x_{k-1}^2 - y_{k-1}^2 + x_k^2 + y_k^2 \end{bmatrix} \tag{16}$$

and x=[x0, y0]T. The location of the target sensor is then obtained by means of solving the least squares problem

$$\min_x |\mathrm{Ax} - \mathrm{b}|^2 \quad 0 \le x_0 \le D \text{ and } 0 \le y_0 \le D \tag{17}$$

In Fig.5, it compares the performance of the triangulation-based localization with the schemes employing distributed learning, It shows that, as the number of beacon nodes increase, distributed learning outperforms the schemes based on triangulation. This is because the triangulation method relies largely on the accuracy of the distance estimate obtained with DHL and DV. The estimation accuracy is affected by many factors such as node density, transmission range, non-uniformity etc. On the other hand, with distributed learning, each beacon node matches the features or characteristics of the non-beacon node to its own in order to compute the local decision.

Conclusion

In this paper, several distributed learning approaces are proposed to localization in wireless snesor networks with only hop-count information at each node. In the system, it is assumed that the existence of a number of beacons that have perfect knowledge of their own coordinates exist and that used their knowledge as training data for distributed learing on the target sensor's location. A series of binary classification problems were solved to determine which area the target sensors falls into. The features that were used to perform the classification were the hop-counts or the hop-distances estimated via the DHL or DV methods. The three approaches, namely HC, DHL, and DV method, were compared under various system parameters. The proposed schemes were also compared to the often employed triangulation method. The simulation results show that the localization methods based on the distributed learning is more accurate and more effectual.

REFERENCES

A. Ribeiro and G.B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks", IEEE Trans. on signal processing, vol.54, Mar. 2006.

D. Zhao, Y. Men, and L. Zhao, "A hop-distance algorithm for self-localization of wireless sensor networks", IEEE ACIS., 2007.

G. Mao, B. Fidan, and B.D.O. Anderson, "Wireless sensor network localization techniques", ELSEVIER journal paper, Computer Networks, vol.43, pp. 499-518, 2003.

Joel B. Predd, Sanjeev R. Kulkarni, and H. Vincent Poor, "Consistency in models for distributed learning under communication constraints", IEEE Transactions on information theory, vol.52, pp. 52-63, Jan. 2006.

Joel B. Predd, Sanjeev R. Kulkarni, and H. Vincent Poor, "Distributed learning in wireless sensor networks", IEEE signal processing magazine, pp. 56-69, July 2006.

L. Doherty, L.E. Ghaoui, and K.S.J. Pister, "Convex position estimation in wireless sensor networks", Proc. IEEE INFOCOM'01,Apr.2001

Luc Devroye, A probabilistic theory of pattern recognition, New York: Springer-Verlag, 1996.

M. Gastpar, M. Vetterli, and P.L. Dragotti, "Sensing reality

and communicating bits: A dangerous liaison", IEEE signal processing magazine, vol.22, pp. 70-83, June 2006.

P.K. Varshney, "Distributed Detection and Data Fusion". New York: Springer-verlag, 1997.

R.Viswanathan and P.K. Varshney, "Distributed detection with multiple sensor: Part 1-Fundamentals", Processing of the IEEE, vol.85, pp.54-63, Jan. 1997.

T.S. Han, S. Amari, "Statistical inference under multiterminal data compresion", IEEE Trans. Inform. Theory, vol.44, pp.2300-2324, 1998.

Y. Shang, W. Ruml, Y. Zhang, and M.P.J. Fromherz, "Localization from mere connectivity", Proc. IEEE ACM MobiHoc, 2003.